A comparative study on the effects of AIM on oral language production skills

Jans, Daan

A comparative study on the effects of AIM on oral language production skills

MA thesis Applied Linguistics
University of Groningen

Groningen, 29 juni 2011

Daan Jans

University of Groningen
Student number: 1458744
Supervisor: dr. M.H. Verspoor

Nieuwe Blekerstraat 12d
9718 EJ Groningen
06 - 385 19 027
d.s.jans@student.rug.nl
daanjans@hotmail.com

ABSTRACT

A comparative study on the effects of AIM on oral language production skills

Nowadays, teaching French as a second language on secondary high schools demands a lot of effort from teachers and the results are often disappointing. It is for those reasons that Wendy Maxwell from Canada developed the Accelerative Integrated Method (AIM), which has become very much used, not only in Canada but also, for example, in Holland.

Though increasingly used, so far there is only little scientific evidence supporting the implementation of AIM on a larger scale. The research that we have performed tries to fill that gap. Whilst my colleagues focus on written proficiency and dyslectic students, I have examined the effects of AIM on oral proficiency.

Using the SOPA method, 4 classes of first year high school students (ages 12-13) have been interviewed and their oral proficiency level has been rated on three different aspects: oral fluency, oral vocabulary and listening comprehension. Two of the classes had been AIM-instructed during one year (56 participants), whereas the other 2 classes (48 participants) had received a more traditional instruction (Carte d'Orange). The test consisted of three different tasks after which they were assessed on fluency, vocabulary and comprehension.

The test results indicate a clear difference between the two groups in favor of the AIM-instructed classes on all three measured aspects of oral proficiency, hence supporting the implementation of AIM on a larger scale.

In addition we looked at the impact of scholastic aptitude on the effectiveness of AIM (no significant effects) and the role of the instructor, where we found some interesting phenomena.

ACKNOWLEDGEMENTS

This research is part of a bigger project in which a multitude of persons and instances are involved. Briefly, I would like to dedicate some words to those people that deserve to be mentioned particularly, because of the assistance they have provided. These people have all been very positive influences and have all contributed in their own way to the realization of this thesis. Even more so, I strongly doubt this thesis would have even been written without some of these people.

I would like to start by expressing my gratitude to the H.N. Werkman College, Groningen. This research took place on their initiative, but during the test we were provided with all necessary means and help. The school was most hospitable in welcoming us during the week of testing, and any assistance we needed was provided as soon as possible. In this respect, some extra attention should go to the two teachers whose classes were subject to the tests, Hanneke Hink and Barbara Koerselman. As we will see in the following, they fulfilled some important tasks in the preparation of the experiment. I felt Hanneke and Barbara were very much committed to the research, which led to a successful co-operation. Any researcher will agree that good circumstances for gathering empirical data can make the research so much nicer and easier. Thanks to H.N. Werkman College for a very smooth week of testing and for supplying an interesting research topic. I hope this research satisfies your expectancies and that you will benefit from it.

Secondly, I would like to express my gratitude toward the two other researchers that have worked on this project. Yvonne Wever, who studied the influence of AIM on dyslectic students and helped during the interviews, has been very helpful. I look forward to hearing about your research. And of course Audrey Rousse-Malpat, who studied the impact of AIM on written language production skills. She was very helpful during the interviews, and almost all of the other parts of this project, and co-operating with Audrey has been a pleasure. She provided serious feedback at the right time, but we also had a lot of fun during the project. So, merci énormément à toi!

I would also like to thank drs. M. van Es for his help with the statistical analyses that were done for this research.

Finally, I want to thank my thesis guide dr. M.H. Verspoor. Anything I could wish and expect from my thesis guide, she delivered. It has been an honor to be assisted by such an experienced researcher and every time we met, I left your office feeling good and confident about my thesis. Thanks for your patience with me (which definitely is a lot!) and your faith.

This thesis is to be the end of 7 years of student-life, and during all of these years my parents have been there for me. I have always felt your support and I know you will keep on supporting me forever.

Thank you.

Daan

INDEX

1 INTRODUCTION

In the Netherlands, French is one of the languages that every student learns in High School. At least, that is the idea. All students have classes of French for some time, but these classes are not always as successful as hoped.

I know from my own experience, that even after 6 years of French in High School, students are not very proficient in French. I found it very hard to express myself in French and found it frustrating to notice what those years of High School had brought me. Note that, if I may say so, I belonged to the upper half of the class in French and was very motivated to learn the language. I participated in a student exchange with a French School in Voiron and ended up choosing to study French at the University of Groningen, where I was told immediately to forget the rules I had learned in High School, and start all over again.

In addition to not achieving very satisfying results, French as a High School subject is not very attractive. It is probably among the most hated subjects and even if France is one the most popular holiday destinations, the number of students that choose to study French decreases every year. Some students just do not like languages, others prefer more 'exotic' languages as Spanish, or languages that are more similar to Dutch (German, English).

Not surprising it is that the Dutch educational institutions are looking for changes that could possible improve the status of French as a subject in High School.

Autumn 2009, we received an email stating that the H.N. Werkman College (one of the local High Schools) was looking for students that were interested in doing research on a method of teaching French they had recently implemented in the first year's classes. At the time I was on the look-out for an interesting research topic for my MA thesis, and this project instantly got me enthusiastic. I was content to see initiatives were taken to improve the subject of French, and the idea of doing research that would actually serve a practical use appealed to me. Apparently, I was not the only one, because soon after I was joined by Yvonne and Audrey, who were equally eager in investigating this new method.

The method in question appeared to be of Canadian origin, which is not illogical considering the Anglo-French bilingual situation in Canada. Our very first theoretical research pointed out that this method had been developed by a teacher of French, Wendy Maxwell, from a dissatisfaction with the available methods of teaching French. Apparently, the Netherlands are not the only country where French as a subject poses problems. Maxwell decided to develop her own method, based on her own teaching experiences and existing research on Second Language Acquisition (SLA), which resulted in the Accelerated Integrative Method (AIM)[1].

AIM is a highly communicative method (more on this later) that presents some very innovative characteristics that are relatively new to second language education. Where the Dutch education of French relies mostly on more traditional methods using handbooks with exercises and wordlists prioritizing the development of written production and comprehension skills, AIM forwards a different approach to the instruction of French.

Two of the most salient aspects of AIM are that from the very first lesson on students and teachers only communicate in French and that this communication is supported by the use of gestures. Thus, the students are presented with a very large amount of input in the target language. Moreover, this input is

[1] Maxwell (2001)

selected by frequency. Only the most frequent (and therefore most important) words and structures are included, making the input useful and valuable. The students learn the most important words first, allowing them to start communicating in French as quickly as possible (thus, Accelerated Integrated Method).

Another eye-catching AIM characteristic is the fact that in the beginning no attention whatsoever is attributed to the development of written language skills. It is only after some time, and with a slow build-up, that AIM involves writing and reading French. Same story for (explicit) grammar. No rules of grammar are given, imminent exceptions withheld. Although this last point has been forwarded by many researchers (Krashen, 1981; Schwartz, 1993), both parents and teachers have expressed their fear this may result in a lower written proficiency. First impressions about AIM have been, in majority, positive, the novelty of the method implies that true academic proof that can take away this fear is scarce.

As already stated, during the AIM classes the only language that is spoken, is French. This does not mean that the teacher constantly has the parole. On the contrary, the students are encouraged to engage in communication from the beginning, resulting in a considerable amount of interaction among students and with the teacher.

Naturally, other methods also claim to encourage communication in the target language, but have mostly failed to do so. Communicating in French may be scary at first, but Aim takes that anxiety away very fast. In addition it stimulates an enthusiastic attitude towards the language by presenting exercises they like to do. Through music, dance, stories and drama, combined with the initial absence of homework, the student's attitude towards the subject becomes more positive. As a result, they become more motivated, which is one of the most important predictors of successful language learning (Gardner, 1985).

In summary, AIM emphasizes - especially in the beginning - the development of oral proficiency. The students receive a lot of rich input, whilst no explicit rules of grammar are taught. AIM is highly interactive and can be considered a communicative method.

The H.N. Werkman College faced the problems with French as a subject as described above. In addition, it was hard to use traditional methods in their heterogeneous classes (classes with students of mixed levels of scholastic aptitude). As a trial, the school decided to implement AIM into 2 first year's classes. This method, being as innovative as it is, raised doubts. In order to obtain a clearer image of the up- and downsides of AIM, we were requested to examine the effect of AIM on the development of French proficiency, and to compare it to the other method they use (Carte Orange, editor: Meulenhoff).

A request that, finally, ended up in this study.

In the next chapter, we will discuss the theories that are related to this method, and the related empirical studies that have been effectuated. Do they support the use of AIM, or do they predict AIM not to be very effective? Following, we will present the set-up of our study and explain the methodological issues we faced and decisions we took, where after the results section lists the outcomes of the study, which are interpreted and explained in the discussion.

2 THEORETICAL BACKGROUND

As stated in the introductory section above, the Accelerative Integrated Method (AIM) has become widely used for teaching French as a second language. Recent years, the originally Canadian method has expanded overseas to countries such as the Netherlands. This popularity can be explained by a number of theories that support the implementation of this method into current SLA programs. As we will demonstrate in this chapter, AIM relies on the support of numerous theories from the domain of applied linguistics that have emerged over the years. We will see that the ideas behind AIM show a lot of correspondences with the communicative approach.

Furthermore, we will examine what research exists on AIM itself and the constructs that it represents. The subsequent question that we will try to answer in this chapter will be whether or not the linguistic research and the modern ideas on SLA lend support to the usage of AIM.

Another question that we want to answer in this paper is under what conditions AIM achieves optimal results. This will, amongst others, be done by looking at the teacher and the specific characteristics that suit teaching AIM best. In addition, a closer look into individual differences and learner characteristics may give an idea of which type of students take most profit from AIM. In this chapter we will examine what research exist up till now regarding these subjects, thus helping us create an idea of what to expect in the following chapters of this paper.

The communicative approach

For centuries, teachers and researchers have worked on finding the most effective way to teach and learn second and foreign languages. Each new theoretical insight on language learning inspired a new approach or method to teach languages.

In behaviorist approaches, the assumption was that repetition and habit-formation were essential to learn a language. Learning took place solely through imitation of conversations and focus on grammatical rules that were intensively practiced and repeated until it became a habit but translating or audio-lingual methods became very fast unsatisfying, even if we cannot deny that these methods had an effect on learning a second language.

People started to lose interest, for one thing because they realized that they were not able to communicate in the second language. Second of all, it corresponded to the time when Chomsky (1966) proposed a new theory that stated that people were able to create sentences and generate patterns endlessly, an assumption that was not in line with behaviorism.

At the end of the 20th century the “Communicative Approach” or “Communicative Language Teaching” (CLT) appeared as the next step in language teaching.

At about that time, Canada decided to seriously work on finding effective L2 teaching methods and started immersion programs using the L2 as instruction medium in the classrooms based on Communicative Learning Theory (CLT). The underlying assumption of CLT is that language is a social activity and that learners should be able to communicate in the target language. The message is more important than the form and the role of interaction is stressed.

This principle goes along with other theoretical claims on second language learning, that, as we will see, can also be attributed to AIM.

CLT stresses mostly on input and particularly on what kind of input should be addressed to learners. It is believed that input has to be authentic but at the same time adapted to the learner's level; the features must be salient and comprehensible.

This focus on meaningful input is the basis of the organization in the classes. L2 instruction is given through activities promoting frequent interaction among the learners obliging them to help each other solve encountered problems. Proponent beliefs in authentic material and real-life situations as well as the relevance of the learner's background are key notions of those methods. According to CLT principles, teachers should have a role of suppliers of relevant input and grammar learning should be inductive. We find these assumptions back in more recent works that have been done in the field of language instruction, in particular in Long's (1991) notion of "focus-on-form".

In Long (1991), "focus-on-form" instruction is defined as follows: (In form-focused instruction) "lessons that focus on meaning are purely communicative [...]. Learners are presented with comprehensible, holistic samples of communicative second language use." (p. 183)

This is opposed to the more traditional "focus-on forms", where "learners are encouraged to master each linguistic item" (Long, 1991; p181). As Long mentions in his article, this type of instruction focuses mainly on the mastering of grammatical rules. Learners talk about the L2 but not really in the L2. This type of instruction tends to be rather rigid. If the linguistic knowledge of the learners improves unquestionably, one could doubt of the benefits of a "focus-on-forms" method in a simple one-on-one conversation.

Most researchers agree on the rather ineffectiveness of focus-on-forms instruction, but debate still remains around the instruction of grammar in form-focused instruction. Some believe that it should be learned explicity whereas others are of the opinion that language acquisition would benefit the best from implicit grammar instruction.

The question is thus tackled differently in each CLT methods. AIM is very clear on that matter: no explicit focus on grammar will be paid in class, however from time to time, some constructions can be supported by a gesture such as word order for instance.

On other theoretical aspects, AIM was set up with CLT theoretical insights in mind. Focus is put on meaningful L2 input which is an absolute key principle of the method. AIM aims at enhancing communication focusing on oral skills. Students begin with a real immersion in the L2 environment as they are taught with a high level of L2 input. Later, they are asked to produce speech in the L2 only, which provides a high amount of interaction and output.

The focus is on a high-level of fluency in oral production and the consideration of second language learning as a means to communicate rather than an object of study makes AIM a CLT method.

Empirical studies

Roughly, we can consider AIM to be a method of communicative SLA, and a lot of its ideas have been inspired by the communicative approach. However, it also draws on principles that are less directly related to the communicative approach. Here below follows a summary of empirical findings and theories that have influenced the design of AIM.

Generally, empirical studies investigating communicative activities, concluded that they had positive effects on learning. They lead to higher accuracy in speaking and writing (Allen, 1989; Spada and Frohlich, 1995) and optimalize learning (Wesche,1994). Besides, motivation to understand increases when learners are involved in such activities.

The past decades have seen a great amount of studies paying closer examination to this distinction between implicit and explicit focus-on-form (Long, 1991). According to Long, implicit focus-on-form occurs only in a meaning and communication-based setting with attention on form. Harley & Swain (1984) however showed that although learners achieved a high level of fluency in this "natural approach", they failed to master some French grammatical features, maybe because of fossilization related to a lack of error-correction.

The discussion about the effectiveness of the two focus-on-form instructions is still a lively one, and no definite answers have been found so far on this topic. Our study is interesting in this debate, since it compares two methods of instruction, one with implicit focus-on-form (AIM) and one with explciit focus-on-form (Carte Orange).

As already mentioned above, choosing the kind of input a method should provide is a process that should not be taken lightly. It is important to have authentic, salient and comprehensible input that is adapted to the learners' proficiency level. But, following Gardner (1985) and his Theory of Multiple Intelligences each students requires different kinds of input according to which learning strategies suit them best and which method stimulates their capacities most. The drama, music and gestures that are very present in AIM serve to ensure that not only the students with high scholastic aptitude, but also the students that present highly developed musical, kinesthetic or interpersonal intelligence, can benefit from this method.

Moreover, AIM follows a content-based structuration of the instruction. The classes evolve around different stories, starting by learning the story's plot and ending by rehearsing a drama-play around that story. Studies (Cummins, 1996; Piaget, 1956) have shown that this content-based instruction provides a furtile context for SLA because it helps contextualize the language perception and production.

In the introduction, we stated that AIM was designed from a wish to break with the existing educational programs and provide a new method of teaching French. Although many studies (Burstall, 1968, 1970; Carroll, 1975) have demonstrated the importance of instructing the L2 in the target language, as Turnbull (1999) shows, teachers of French still rely heavily on the L1, speaking to up to 63% of the time in the L1. In AIM, the objective is to speak the target language (supported by gestures) as much as possible, thus maximizing comprehension and oral production skills of the students.

When it comes to the choice of the vocabulary input that AIM provides, it also ruptures with existing methods. Inspired by students' frustration of not being capable of engaging in the most simple communicative activities, AIM has restrained the vocabulary inserted in the method to only the most basic (but important) vocabulary needed for communication, in order to develop communicative skills as quickly as possble. This vocabulary has been selected according to frequency and function.

The importance and advantages of learning highly frequent words first has been shown by various researchers (Barry & Seymour, 1988). They constitute the most important linguistic features around which the rest of the language is built up. Compared to other methods, AIM goes very far in emphasizing these words and neglecting other less frequent words.

In mastering these basics of French, repetition is an important means for the students to fully acquire the meaning and usage of this selection of essential words. Fewer words, means more repetition, means more intake of the targeted word, according to Sharwood Smith (1993). Repetition enhances the richness of the input and is one of the most effective means to stimulate intake of linguistic features. Menezes (2009) confirms this claim about the importance of repetition stating that "language learning is

understood as the processing of experience and the repetition of experiences causing the strengthening of the connections"[2].

In summary, as Wendy Maxwell (the developer of AIM) describes, the method "is designed to support comprehension and production through a high degree of repetition of essential vocabulary, embedded in context and experienced kinesthetically and visually"[3].

Studies on AIM

Because of AIM's relative novetly in the field, the empirical studies available are scarce. The following section gives an overview of research to date on AIM.

Several studies on AIM have been conducted, mainly in Canada between 2001 and 2009. The largest part of those studies focuses on oral production, but are different to our study in the number of participants or in methodology.

Maxwell (2001) compared the oral fluency of two groups of 9 students (AIM/ non-AIM), who were interviewed with a scaffolding questionnaire and who were asked to create a story spontaneously.

Her results show that AIM students outperformed non-AIM students, but lacked the statistical evidence due to the limited number of participants. Quantitative results on inter-group interviews pointed out that AIM students of different aptitude levels performed more homogenously during the interview than non-AIM students.

According to Maxwell, "the results are interesting in that they indicate that this type of approach responds to the needs of a variety of the students and that the average learner may thrive as well or better than the academically strong"[4].

Interestingly, Michels (2008) obtained comparable results in his replication study. However, it may be difficult to generalize these findings because they both had a very limited number of participants.

Although lager scale studies with statistical analyses have been conducted on AIM, none have corroborated a significant difference in French proficiency between AIM and non-AIM students.

Mady, Arnott and Lapkin (2007, 2009) compared six classes of 13 year-old grade 8 AIM (n= 125) with 6 classes of non-AIM (n=135). Using a mixed-method study composed by a test-package for proficiency (Harley, Lapkin, Scane, Hart & Trépanier, 1988) and a questionnaire on perception of French classes, they concluded that there were no significant differences between their language skills and their perception of French as a L2. However, on a qualitative level they found a major discrepancy in the perceived factor believed to be the key to success in the L2. Non-AIM students attributed it to the teacher, whereas AIM students pointed out the method. Asked on their perceived development of the L2, AIM students answered that they felt "better than before" but their comments on writing skills were mostly negative.

A follow-up survey revealed that, one year later, the continuation rate of AIM and non-AIM students was similar.

[2] Menezes (2009), pp. 9

[3] Maxwell (2001), pp. 2

[4] Maxwell (2001), pp. 36

In Bourdages and Vignola (2009), results show no significant differences in linguistic or grammatical accuracy between AIM and non-AIM students. However, they noticed that AIM students seemed to have a wider vocabulary and that they talked significantly more French.

In Arnott (2005), this difference in attitude was further investigated, particularly the amount of risk-taking that AIM students dared to take compared to non-AIM students. Students said during their interview that they were able to handle a French-environment.

Clearly, mixed results have been found concerning the potential benefits of the AIM on linguistic proficiency. The only clear statement that can be made on AIM according to research to date is that AIM students deal differently with their L2 possibly due to an enhancement of creativity and fluency and a willingness to take more risk.

The above clearly shows that AIM may be a very promising method of teaching French. Though still in the process of development, it has already proved its success in Canada and is rapidly expanding to other nations. Teachers and students are enthusiastic, and AIM draws on some established theories on SLA. To date, research on AIM is scarce and not conclusive; a gap that this study tries to fill. In addition research has so far focused on the oral production using small groups and questionnaires. Our study is big enough to justify statistical analyses and will presents a different methodology, on which we will elaborate in the next chapter.

The main research question of this paper is:

Do students who have been taught by AIM have better oral language production skills than the students of the control group?

In addition we want to examine if other factors play a role in the effectiveness of AIM. As we have seen, one of the objectives of AIM is to appeal to different kinds of students by not only stimulating their linguistic capacities, but also their visual, musical and kinesthetic properties. We thought it would therefore be interesting to investigate whether general scholastic aptitude of the individual student has an effect on the impact of AIM on their language production skills. Hence, the formulation of a second research question of this study:

Does the effect of AIM on the oral production skills of the students depend on their initial scholastic aptitude level?

A last relevant topic on which we wish to shed some light in this paper is how AIM is impacted upon by the teacher. Do all teachers acquire the same results using this method, or can we find different results for different teachers? Hopefully this study will find an answer to this third (and last) research question:

Does the effect of AIM on the oral production skills of the students depend on the teacher?

3 METHODOLOGY

In the paragraphs below, we will present all aspects concerning the execution and content of the experiment, and we will explain the reasons behind the decisions that were made in preparing and executing the tests.

1. Subjects

Since the implementation of AIM on the H.N. Werkman College was still in an early phase at the moment of testing (it was not even sure whether the school would continue AIM instruction in the year after), not all first year classes were taught French using AIM. In fact, the school had chosen to start AIM courses only in two first year classes. Depending on the impact of AIM on students and teachers, the school would make decisions about any further use of AIM. Apparently, the effects of AIM satisfied the school, and AIM was maintained and increased the year after.

This did imply that no more than two AIM instructed classes were at our disposition for this research. Dutch high schoolclasses generally contain about 30 students, and this held for the AIM classes as well. In order to examine the effect of the AIM instruction on the results we compared those two AIM classes to two other first year classes that received instruction following a more traditional method of learning French: *Carte Orange* by Thieme Meulenhoff.

We were in the fortunate position that the two teachers that instructed the two AIM classes, instructed two first year's Carte Orange classes as well, which we could include in the research as control groups. In total, we tested 4 first year classes, including two classes that received AIM instruction (thus forming the experimental group) and two Carte Orange classes (the control group).

In the end, we managed to gather a sample of N=45 for the control group, and N=50 for the experimental group. We had some fall out among the subjects due to illness, but we had a total number of 95 participants.

The groups can be considered comparable in many different aspects. For one, they were all first year high school students aging from 11 to 12. There might have been some outliers that have skipped or redone one school year, but the claim can be made that the factor age was similar for all groups and did not interfere in the results of this study.

In addition, the participants shared relatively similar backgrounds. Each participant regarded Dutch to be their mother tongue, and up till the moment of testing their educational history was comparable. None of the students had learned French before, even though it is likely that some of the students had already received some input of French through holidays, media, etc.

In The Netherlands, all students take a test in the final year of elementary school, called the CITO-test. The scores on this test are seen as indicator of intelligence and scholastic aptitude of each student, and generally, High Schools divide their students into streams of different ability according to their CITO-scores. However, some schools choose not to distinguish students according to their scholastic aptitude during the first years, and postpone this selection process to a later stage. The H.N. Werkman College is one of the institutions that does so. This means that during the first two years of high school, no distinction is made between students according to their capacities. Thus, they are placed in heterogeneous classes: a 'melting pot' of students of different levels.

This leaves us researchers with both advantages and disadvantages. On the one hand, this means that our study includes students of different aptitude levels. If done properly and cautiously, the influence of aptitude on the effect of AIM can be inserted as a research question.

However, such a divergent group of students can also make it more difficult to draw valid conclusions about the results. In order to generalize the findings we need to make sure that the groups are as equivalent as possible. We gathered the CITO-scores of most of the students (92 out of the 95) which we used to compare the two groups. An Independent Samples T-test was carried out to find out whether or not the experimental and the control group have significantly different average CITO-scores.

No statistically significant difference was found between the CITO-scores of the AIM instructed students (M=536.89, SD=6.190) and the CITO-scores of the Carte Orange instructed group (M=537.8, SD=7.099).

Therefore, we can consider the groups to be comparable when it concerns their aptitude levels.

We feel that using the CITO-scores was a valid option to measure aptitude in this study. During the last year of elementary school all pupils take the final CITO-test. It measures the different variables involved in learning processes, thus fulfilling its primary function as predictor of future educational success. The test establishes how successful a learner the pupils have been during their time at elementary school, which includes numerous other factors besides intelligence and aptitude. However, as we have seen in the previous chapter, these last two variables are very dominant in predicting scholastic performance, and therefore also largely determine the outcome of the CITO-test.

Another important factor that determines the success of the education is the instruction itself. As mentioned above, the four classes were instructed by two different teachers, each teacher instructing one AIM class and one Carte Orange class. No teacher teaches in the same way and that holds for the two teachers involved in this study as well. Comparing the results of the different teachers may also be very interesting. Yet, in comparing the experimental to the control group we need to be careful about assuming the teachers are the same. Both groups had one class instructed by each teacher. The teachers used the same manual, discussed the program to follow, and the students were provided with an equivalent amount of exposure, both quantitative and qualitative. Therefore, we can assume all students were provided with the same information, and any possible within-group difference was sorted out by averaging both classes in each group and looking at the two groups as one whole.

A final remark, before we start looking at the test procedure itself, is about the amount of input. The students of the H.N. Werkman College had 3 classes of French a week, which is considerable for first years students in The Netherlands. Each class takes 45 minutes, so the students had 135 minutes of French contact hours a week. The students were tested after the first school year. The number of contact hours is estimated to be 27,75 hours, or 1665 minutes, based on 37 schoolweeks a year. In addition to the contact hours, the students were regularly asked to do some homework. This homework consisted of doing exercises, learning vocabulary, etc., and preparing for tests. One major reason why the students indicated that they liked AIM very much is that this method starts giving homework and testing a lot later than Carte Orange. During the first months, the AIM students were not asked to do anything outside of the classroom. This made the method very popular amongst students, but it also implies that it is very likely that the total amount of hours contributed to learning French was a bit less for AIM students than for the Carte Orange students.

2. Testing procedures

Testing for oral proficiency is time consuming and demands a lot of effort from both researcher and participant. In addition, oral proficiency is hard to test for and rate. Therefore, designing an oral

proficiency test demands extensive reflection and preparation. In this chapter, we will present the outline and execution of our oral proficiency test, and simultaneously justify the choices we have made along the way.

Let us start by signaling one of the downsides of this study. Considering the effort needed to orally test 95 students and the limits of this research, we were only able to test the students once. This allows to make comparisons between the groups, but unfortunately, it does not tell us anything about the development over time. Testing oral proficiency development over time could indicate how the groups progress, and might result in interesting conclusions. Future research could include the data that will be presented in this paper in a longitudinal study to look at the development of AIM students over time. Already, new projects examining AIM have been set in motion and I would consider the further oral development of AIM an interesting research topic.

Finding the right method of measuring oral proficiency was a major challenge. Since the students were still relatively young of age (11-12 years) and only received French instruction for almost one school year, we could not expect the subject to have a fluent conversation in the target language. In addition, looking back at our own high school experiences, we were concerned that students would not be very motivated, or even, embarrassed or frightened to speak French.

Then again, we were very enthusiastic to discover a method of oral proficiency testing called Student Oral Proficiency Assessment (SOPA). This method was developed in 1991 by the Center for Applied Linguistics (CAL) for Spanish students of English of about the same age as our subjects. However, over the years, the method has been increasingly used to test students with other backgrounds as well. SOPA is, as we will demonstrate later on, a method that is age-specific. Even though the content of the assessment can be adjusted to the target group, the test can quickly become too childish for some subjects, or too demanding for younger children.

The main objective of SOPA is to establish what the subjects *can* do, instead of what they cannot do. The interview-based format that SOPA uses aims at eliciting the highest proficiency level the students can achieve. The concerns we mentioned in the section above are shared by the developers of SOPA. Speaking in a new language can be frightening to kids around 11-12 years of age and can have such an impact on them that they do not perform as well as they could have. This 'negative washback' is something to avoid because it would give misleading results.

To prevent this 'negative washback' from occurring, SOPA's main concern is to always make sure the subject feels *at ease* during the interview. The underlying thought is that the students will only demonstrate their full capacities in a situation where they feel comfortable. This 'comfort zone' that the students have to be in during the interview implies that the students should not feel pressured perform. The entire outline of SOPA is guided towards finding the students' highest possible proficiency level without causing the student to leave its 'comfort zone'. According to CAL, the ideal context for high oral performance for students in this age category is a friendly and non-stressful environment.

Setting

Before we start looking at the content of the SOPA design, we will first explain the practical setting in which the experiment took place. SOPA is designed in such a way that 2 students can be interviewed and rated at the same time, but the interviews do require the active presence of two researchers. Thus, the interviews are held in a setting where the two students sit opposite the two researchers.

One of the researchers is the interviewer and is the one that leads the students through the different tasks and keeps the conversation going. We will come back later to the exact role and responsibilities of the interviewer.

The other researcher is responsible for the ratings. During the interview, s/he fills in the rating sheets that contain some practical information (names, class, date, etc.), but whose main purpose is to help decide on the final assessment of the students' performance. The rating sheet contains a separate box for each task in which remarks and notes can be written down that are relevant to assessing the oral proficiency. Appendix I contains an example the rating sheets that we used for this study.

Each interview took about 20 minutes, followed by a 10 minute break in which the two researchers had to come to an agreement on the assessment of the students. More on this assessment later. These 10 minutes were usually sufficient to rate the students and to prepare for the next interview.

Another responsibility of the researcher who did not perform the interviews was to operate the video registration equipment. No matter how consistent and objective raters try to be in rating the students, they remain subjective to a degree. To avoid inconsistencies in the ratings all of the interviews were recorded to verify the scores. After the experiment, the raters went through the recorded interviews to ensure a consistent and correct assessment of the students.

Admittedly, the fact that the students were recorded with a video-camera did not decrease the stressfulness of the situation for the students. Looking at some of the videos, you can see some of the students being startled by the camera. Especially at the beginning of the interview, some of the students keep looking directly in the direction of the camera. However, as the interview proceeded, they mostly forgot the camera was turned on, and as long as they were forced to actively participate in the interview, the students' performance seemed not to be affected by any video-camera related stress. Then again, part of the introduction preceding the interview was to emphasize that the video registration would only be used for academic purposes and to verify if we had proceeded correctly. After this explanation, we asked the students if it was okay to record the interview. None of them objected. Since we had asked the teachers to prepare the students by saying - amongst others - that the interview would be recorded, the video-camera was no surprise to most of the students. The students may well have been aware of the video-camera's presence, but it did not bring them out of their 'comfort zone'.

The interviews took place from Tuesday the 8th of June until Monday the 14th (Saturday and Sunday excluded). During those 5 days we interviewed all of the 95 students. Each day we were scheduled to interview between 18 and 22 students, but unfortunately we had to deal with some drop-out due to illness. The schedule was made by the teachers, who also informed the students where to be at what time. This worked out perfectly. The students were granted permission by the school to leave their classes (the other teachers were informed as well) when they were supposed to do the interview. In general, the students were surprisingly faithful to this schedule, and we only had to go search and take the students out of their lessons twice because they had forgotten the interview.

Another important reason we asked the teachers to make the schedule, was because they were most capable of assembling the right pairs of students. To avoid situations where the one student would dominate the conversation over the other, we needed to make sure the difference in proficiency of the students that would participate in interviews together would not be too big. In addition, we asked the teachers to avoid pairing very introvert students with more outgoing students.

The school provided us with a classroom that we could use during the week of testing. Speaking in another language requires a lot of concentration, and it was of great importance that the students were not distracted by any surrounding noise whatsoever. Plus, the circumstances had to be equivalent for all students. We were very content to find that the classroom that we were appointed was situated in one of the back corners of the building, which was relatively quiet. In addition, during the week of the interviews the higher grades had a week of exams. This means that they did not attend classes, and only

came to school to do a few tests. The classroom was peaceful and provided us with an excellent setting for our interviews.

Overall, we were very satisfied by the context we created for the interviews and by the rest of the practical organization. In the next paragraphs we will go into the content of the interviews and the underlying ideas.

'Comfort zone'

As already mentioned, SOPA's design aims at establishing the students' highest proficiency level which can only be found within a friendly and non-stressful setting. In this section we will go over the different strategies that contribute to the objective of making sure the students stay within their 'comfort zone'.

The most important aspect in this regard is to start off with simple tasks that the students can easily execute, followed by tasks that increase gradually in difficulty. In fact, the first tasks are designed in such a way that they should be easy for all students. This serves as a good manner of breaking the ice and it gives the students confidence. However, the students have to maintain focus, and the tasks have to remain challenging. Here lies a great responsibility for the interviewer. We found that there were some great differences in proficiency level amongst the students, so the interviews (and the level of difficulty) had to be adapted to the students. When it was clear that a student was capable of performing a task, it was time to move to the next, or to increase the difficulty. But, in the case where a task appears difficult to a student, the interviewer has to help the student or decrease the level of difficulty. This process can go on and on until it is clear that the student cannot cope anymore with a higher level difficulty. When the student's *ceiling level* has been found, it is time to start taking down the difficulty again so that the student can feel positive about his ability level and wind-down the interview. The highest level we found is the rating that we used to assess the students with, thus looking at what the student was capable of, ignoring what he was not capable of.

This process of starting off slowly and gradually increasing the difficulty can also be found in the tasks themselves. We will come back to exact tasks later, but for now, it is relevant to note that the tasks are designed so that the follow-up of capacities required for the tasks follow the same development as a natural order of acquisition: receptive before productive tasks. The first task focuses on receptive skills (listening comprehension), followed by increasingly difficult tasks that focus on productive skills (oral production).

In making the students feel comfortable, the interviewer plays a great role. S/he is responsible for the all important atmosphere in which the student have to perform. Before the interview sets off, s/he is the one that introduces the researchers to students, but also the one who prepares them for the interview by making sure they feel at ease. Once the interview has started, s/he should always maintain his calm and his friendliness. Where necessary, s/he helps the students. Not by giving answers or completing their sentences, but by posing open questions. For example, if one of the students cannot find something blue, the interviewer can take a green item and ask "Is this blue?" to help. This kind of assistance helps the students a lot, but does not prevent them from showing what they can do.

Another useful and required means to make the students comfortable is to compliment and encourage them at all times. The students should not be afraid to make mistakes and a compliment can boost their confidence.

So far, we have only discussed the students individually. However, at each interview, two students were present. This also served to make the students feel more at ease. They can help each other (peer assistance) and not being alone is less frightening and embarrassing. In addition, the interviewer can direct his questions from one student to the other as he likes. The best interviews were the ones in

which it what like we were just having a talk and playing some games amongst 'friends'. The fact that three people are involved in the conversation makes it more natural and really facilitates the creation of a friendly environment.

At the same time, the danger of interviewing two students lies in one dominating the other, thus preventing the other from showing what he is capable of. Therefore we had to make sure the students would be similar concerning personality and proficiency. If one outperformed the other, that would influence the results, and if one student would just be more talkative than the other, likewise. Since, we did not personally know the students, we gave the teachers instructions for the selection of the couples on the basis of personality and proficiency level. This worked out well.

The last two strategies we employed to make the student feel comfortable have a more psychological character. We mentioned the term 'negative washback' before, but another factor that can impact on the result is 'positive washback'. This means that most of the time that when asked to do a test, students will prepare for it. The higher the stakes, the more preparation, thus influencing the result and blurring the actual proficiency level. This is why we decided not to grade the students and tell them explicitly beforehand. We avoided the term 'test' and only spoke of interviews, the outcomes of which were not counted towards their grades. However, after the interview they did get a reward as a way of saying thanks for their participation: a candy-treat.

Interview

SOPA proposes a number of different interview formats and tasks suitable to different groups and subjects. These tasks (or games) have been designed in such a way that they correspond to the target group and elicit speech in the target language. However, the interview formats as proposed by SOPA did not completely correspond to our needs and/or facilities and had to be adjusted slightly.

Designing and preparing the interview involved different steps. First, we needed to decide on the tasks/games we wanted to incorporate in the interview, and how. Second, we needed a more profound and detailed execution of the interview format, which resulted in an interview script. Third, we had to verify if the interview corresponded to our wishes and practical boundaries. Of course, the theoretical background and SOPA instructions were consulted in the process of decision-making, as we will demonstrate in the coming paragraphs that will elaborate on the interview design process.

Due to time issues – we could not spend more than 30 minutes on each interview. We concluded that we could ask the students to perform only 3 tasks per interview, plus a small wind-down activity. According to the SOPA manual[5] and the slow build-up of level of difficulty we needed to start with an identifying and naming task. The second task should include answering informal questions, and in the third task the students were supposed to follow and give instructions and describe situations. As mentioned in the previous sections, the wind-down activity aimed at bringing down the level of difficulty to make the students feel comfortable and close down the interview.

The identifying and naming task that the SOPA manual proposed was not suitable for our interview because of one major reason. It included naming kinds of fruits and vegetables. Our group of participants had not yet learned a sufficient amount of fruits and vegetables so they could not be presumed to know their names.

Therefore, we had to design another task that the student were presumably capable of executing. We came up with the following task: Naming and identifying different types of playmobil figures. The task was named *Le sac magique* (the magic bag) because the students had to start by opening the bag and

[5] Administrator's Manual For CAL Foreign Language Assessments, Grades K-8 (2009)

taking the figures from the bag. Afterwards, they were asked different question, mainly to test their listening comprehension skills. In some cases, when students needed more challenging tasks, we also asked for descriptions. It appeared that students really liked this task. By letting them 'play' with playmobil, they immediately felt at ease, probably because they noticed that this clearly was not a normal school test, which they obviously liked.

The figures included the different kinds of people (woman, man, boy, girl) wearing different colours and in some cases with some particular features (carrying a briefcase, wearing a hat, etc.). The magic bag also contained a boat, a car, a table and a chair. These figures gave us a very large number of possibilities to test their listening comprehension skills and to differentiate in level of difficulty. The most easy request would be – for example - to point out a woman, but we could also ask for the guy wearing the green hat to be put underneath the table. This task also allowed us to test whether the students were capable of counting in French. In most cases, the students did not seem to have too many problems following our orders, so we could move on quickly to the next task.

The second task was all about eliciting as much speech as possible in the target language. In order to do so, we printed out 6 photo's that lead to 3 different topics (school, weather and sports) we could discuss with the students. Beforehand, we asked the teachers which topics corresponded best to the things the students had learned during their lessons. Hence, we could assume that the students disposed of a certain vocabulary on the different topics. During each interview we talked about two of the topics, each student talking about one photo per topic. In this way we tested to what extent the students were capable of having an informal conversation in the target language.

The last activity proved to be the most difficult as well. The students were supposed to follow and give instructions and describe situations. In order to do so, we presented them with a large depiction of a farm with replaceable stickers that represented the different things one might find at a farm. Again we checked with the teachers if the students could be assumed to know the vocabulary needed to perform this task, which they could. Still, it proved challenging for some of the students to follow orders like "pick up the farmer and place him in the house". Describing where the figures were placed was even harder. Clearly, this task was the most challenging task, and then again this is also the task where we could identify most differences among the students.

The wind-down task involved a game that is commonly known as 'Simon says'. In this case the interviewer played the role of Simon, giving an order to be executed by the students. These orders were very simple: "Stand up", "turn around" and "clap your hands" are examples of the orders we gave in this task. The students often laughed about the things they had to do, which was a positive sign indicating the wind-down activity served its purpose.

Rating

An important, yet very tricky, part of the testing procedures was assessing the students' performance with a score in order to transform the interview results into measurable data. These scores had to be assigned carefully in order to obtain reliable data. Therefore, a number of preparative actions and precautions were taken before the interviews, and the scores were checked in multiple ways afterwards. In this paragraph we will present how we rated the students' performance and the precautions we took to ensure that the data would be as reliable as possible.

During the preparation of the interviews we quickly realized that we were facing a very special group of participants for which no default rating sheet would be sufficient. All of the students had been studying French for only 135 minutes a week during a period of one year, so all of the students could be considered starting learners of French and could not be expected to have obtained an advanced proficiency level (no more than an A1 level in the CEFR). We had to find a way to distinguish different

levels of proficiency within this beginner category. In addition we were facing two subgroups of students who had received very different forms of teaching and thus had learned very different parts of the language. Our question as how to do justice to two such different groups in one rating scale?

We found that the *CAL Oral Proficiency Exam and Student Oral Proficiency Assessment Rating Scale* (COPE/SOPA-RS) corresponded well to our needs. The rating format (appendix II) provides nine scales of oral proficiency for four of the competencies required for oral communication. This rating format, especially designed for COPE and SOPA testing, distinguishes between three main proficiency levels (novice, intermediate and advanced) and each has three sublevels (low, mid and high). In our research we did not use these same names for the ratings. Instead, we just numbered the ratings from 1 to 9, 1 being the lowest and 9 being the highest.

The main reason why we chose this rating format is because it differentiated very well among the lowest levels of proficiency and gave clear examples of the linguistic features that belonged to the different categories. However, the highest possible level (9, advanced-high) was far from achievable for our students as a student was supposed to – for example – "organize and extend discourse (multiple paragraphs) in an emerging ability to hypothesize on abstract topics (if-then) and support opinions"[6]. Clearly, no beginner learner can be expected to be able to meet such demands. Then again, the rating scale exceeded by far the extent of proficiency levels we needed to rate our group of participants. We did not consider this to be a problem for our research. The Rating sheet distinguished between the lowest levels of proficiency in a clear and detailed manner, and that is exactly what we needed. All of the proficiency levels provided by the COPE/SOPA-RS that we did not use, we simply ignored in our research. We ended up with four levels of proficiency (1-4).

As mentioned before the COPE/SOPA-RS rates four different competencies that play a role in oral communication (Oral Fluency, Grammar, Vocabulary and Listening Comprehension). We decided to focus on only three (Oral Fluency, Vocabulary and Listening Comprehension because we felt that students who had only learned French for one year could not be expected yet to have acquired such a proficiency level that rating their grammar in oral production would be an asset to our research. During the interviews this appeared to be a justified decision, because none of the students seemed to pay any attention whatsoever to grammatical features and mostly stuck to using memorized expressions.

To elaborate on how Oral Fluency (OF), Vocabulary (VOC) and Listening Comprehension (LC) have been assessed and what linguistic features typically appertain to the ratings we listed the instructions from the COPE/SOPA-RS and some examples of expressions we found per oral production competency in the following figures. For each of these figures holds that the second row represents the rating that corresponds to the instruction as given by the COPE/SOPA-RS and the examples drawn from the notes taken during the interviews. These have been chosen because we consider them stereotypical examples of students' oral production for that particular rating.

In rating the Oral Fluency of the students we assessed how well the students were capable of formulating sentences and having a conversation in French. In most cases conversations during the interview did not surpass the level of question (interviewer) – answer (student), generally consisting of only one word or expression. Other relevant questions to assessing OF are: How fluent is the students' speech? Does the student try to build sentences or does he use isolated words only? Is the student capable of expressing ideas in French?

[6] COPE/SOPA-RS, CAL (2009), from: Administrator's Manual For CAL Foreign Language Assessments, Grades K-8 (2009)

Oral Fluency			
1	2	3	4
-Produces only isolated words (i.e., single-word responses) and/or greetings and polite expressions such as *good morning* and *thank you*.	-In addition to isolated words, uses phrases of two or more words, and/or memorized phrases or sentences (e.g., *My name is…*, *I don't know*) in predictable topic areas. -May attempt to create sentences, but is not successful. -Long pauses are common.	-Uses memorized expressions with reasonable ease. -Shows emerging signs of creating with the language to communicate ideas. -Creates some sentences successfully, but cannot sustain sentence-level speech.	-Goes beyond memorized expressions to maintain simple conversations at the sentence level by creating with the language, although in a restrictive and reactive manner. -Handles a limited number of everyday social and academic interactions.
"j'aime voetbal" "il est petit" "je fais…"	"je ne sais pas" "c'est jolie" "maintenant? Non."	"il est sur la maison, c'est ici?" "je pense que…"	"je cherche pour un sport" "c'est beaucoup de mots nouvelles"

The assessment of the Vocabulary knowledge of the students consisted mainly of identifying the level of difficulty of the vocabulary and idiomatic expressions used by the students. Some only knew the most basic words and did not say much more than *oui* and *non*. Others used more difficult words and expressions and sometimes knew the context they had to be placed in. Use of different verb conjunctions and tenses also contributed to a higher vocabulary score.

Vocabulary			
1	2	3	4
-Uses single words in very specific topic areas in predictable contexts. -May use greetings and polite expressions.	- Uses single words, short phrases, greetings, polite expressions, and other memorized expressions on a limited number of topics. -Frequent searches for words are common. May use native language or gestures when attempting to create with language.	-Uses vocabulary centering on basic objects, places, and common kinship terms, adequate for minimally elaborating utterances in predictable topic areas. -Use of native language and gestures is common to expand topics.	-Has basic vocabulary for making statements and asking questions to satisfy basic social and academic needs, but not for explaining or elaborating on them. -Use of some native language is common when vocabulary is lacking.

“personnes…marcher” “ça va bien” “ne comprends pas”	“avec le garcon” “manger l’animal” “il y a des nuages”	“ma nom écrit à la papier?” “c’est blanche et noir un petit peu” “le chat est sur la maison”	-

Listening comprehension assessment is based on how well the students understood the interviewers’ intentions and speech. The fourth row of the figure below is empty since we cannot provide concrete examples of understanding.

Listening Comprehension			
1	2	3	4
-Recognizes single, isolated words, greetings and polite expressions.	-Understands predictable questions, statements, and commands in familiar topic areas (with strong contextual support), though at slower than normal rate of speech and/or with repetitions.	-Understands simple questions, statements, and commands in familiar topic areas, and some new sentences with strong contextual support. -May require repetition, slower speech, or rephrasing.	-Understands familiar and new sentence-level questions and commands in a limited number of content areas with strong contextual support for unfamiliar topics. -Follows conversation at a fairly normal rate.
-	-	-	-

After we had prepared the entire interview, and had established the rating scale we decided to do an experimental interview beforehand to test our interview, to practice the rating procedure and to make any necessary adjustments afterwards. We managed to find two suitable test-participants on whom we could test our interviewing and rating skills. This test proved to be very useful and a good training for the actual testing.

The steps en decision explained above are all part of the preparation before the interview. We already explained the role (and the rating sheet s/he uses for that role) of the second researcher present during the interview: taking notes. The note-taker’s task was to write down all linguistic features that were relevant to assessing the students’ performances. Thus, he wrote down general ideas and concrete examples of both positive and negative speech on the part of either student. After the interview, the notes were used as a guideline to assessing the students with scores.

The final measure to ensure a consistent and reliable assessment has already been mentioned as well: the video registration. All of the interviews were recorded and watched again afterwards to verify that the students had been attributed the right score. In general, this appeared to be the case and if not, we adjusted the scores.

4 RESULTS

The interviews, as described in the previous chapter, provided us with some very satisfactory data, as we will demonstrate in the present chapter. But, in addition to the data we retrieved from the rating system we developed, the interviews left us with a great number of ideas, impressions and thoughts about the performance of the students and the impact of the type of instruction on that performance.

Evidently, camera-registered interviews in the target language provides a dataset that is way more rich and subjective of nature than for example the outcomes of a questionnaire. On the one hand, this makes the data a lot harder to interpret. Tests that provide outcomes of numbers on a scale from 1-10 are much easier to handle, because, as the expression goes *numbers do not lie*. A hand-rated interview is more subjective. Although we firmly believe in the objectivity and the reliability of our ratings, other factors can influence (be it subconsciously or not) your decisions and ratings. Consequently, as demonstrated in the previous sections, all possible measures have been taken to ensure proper testing and rating in our research. Therefore, we can be sure about the reliability of our ratings.

However, the interviews are so rich by nature that to only look at the ratings that have been retrieved from the interviews would be a waste of precious data. Therefore, this chapter will not only treat the quantitative outcomes of the interviews (ratings) we have also included a section in which we present a qualitative analysis of the differences in the performance of the groups of students.

Quantitative results

The very first thing we did in analyzing the obtained data, was to control if we were allowed to group the three skills we rated into the bigger main-skill of oral proficiency. If we wanted to proof that the three competencies of Oral Fluency, Oral Vocabulary and Listening Comprehension can actually be seen as measure of one bigger construct, we had to show that, even though we were dealing with three different skills, a lot of correspondences existed between these measures as well. So, in order to verify if all three measure measured, in the end, about the same thing (oral proficiency), we executed a correlation analysis on the ratings of the three competencies, of which the results can be found below.

	Fluency	Vocabulary	Comprehension
Fluency	XXX	0,599	0,612
Vocabulary	0,599	XXX	0,557
Comprehension	0,612	0,557	XXX

Figure 1. Relationships between the three competencies

A Pearson R correlation analysis showed that there was a significant positive relationship between the three identified competencies of oral proficiency. This holds for the correlation between Fluency and Vocabulary (R=.599; p<.05 (two-tailed)), Fluency and Comprehension (R=.612; p<.05 (two-tailed)) and for Comprehension and Vocabulary (R=.557; p<.05 (two-tailed). In all cases this correlation can be considered reasonably strong.

These strong correlations indicate that the three rated skills measure more or less the same things. Due to these relationships we are allowed to group the results of the three measured skills into one larger group, which will hence be called Oral Proficiency.

AIM vs. Carte Orange

Now that we have established that the three measured skills can be grouped into the one construct of Oral Proficiency, we can start looking at the results of the AIM-group versus the control group instructed according to the Carte Orange methods.

In general, the AIM students performed a lot better than their co-students in the control group. In the table below, we can clearly see a big difference in the scores of the two groups (note that the ratings were done on a scale from 1 to 9, though we did not find any score higher than 4).The results shown in the table include the ratings of all three rated skills of all students, which we have averaged down to one single score per group.

	N	Mean	SD
AIM	47	1,461	0,504
Control	45	1,089	0,218

Figure 2. Descriptives of Oral Proficiency results of the two groups

Evidently, a very big difference existed in the performance of the groups, especially when realizing that the lowest obtainable score was 1. The control group score is just above 1, so in general they were very close to scoring the lowest possible rate. The mean score of the AIM obviously lies considerably higher. These impressions are confirmed by the statistics.

An independent samples T-test showed that the AIM-group scored higher (M=1,461; SD=,504) on Oral Proficiency than the Control Group (M=1,089; SD=,218). This difference was significant (t(92)=4,633; p<,05).

Similar tests have been executed for the results of the individual competences, which will now be elaborated.

The descriptives of the test results concerning the Oral Fluency of the two groups are shown in figure 3.

	N	Mean	SD
AIM	47	1,49	0,748
Control	45	1,02	0,149

Figure 3. Descriptives of Oral Fluency results of the two groups

Just like the averaged scores on Oral Proficiency, these data have been tested for significance, providing the following outcomes.

An independent samples T-test showed that the AIM-group scored higher (M=1,49; SD=,748) on Oral Fluency than the Control Group (M=1,02; SD=,149). This difference was significant (t(92)=4,195; p<,05).

The same procedure was followed in analyzing the test results of Oral Vocabulary and Listening Comprehension, of which the descriptive and statistical test results are shown below.

	N	Mean	SD
AIM	47	1,26	0,488
Control	45	1,07	0,252

Figure 4. Descriptives of Oral Vocabulary results of the two groups

An independent samples T-test showed that the AIM-group scored higher (M=1,26; SD=,488) on Oral Vocabulary than the Control Group (M=1,07; SD=,252). This difference was significant (t(92)=2,345; p<,05).

	N	Mean	SD
AIM	47	1,64	0,529
Control	45	1,18	0,387

Figure 5. Descriptives of Listening Comprehension results of the two groups

An independent samples T-test showed that the AIM-group scored higher (M=1,64; SD=,529) on Oral Vocabulary than the Control Group (M=1,18; SD=,387). This difference was significant (t(92)=4,784; p<,05).

In summary, the mean score of the AIM students is significantly better on all rated competences compared to those of the Carte Orange control group. Figure 6 displays a graphic summary of the difference between these competences.

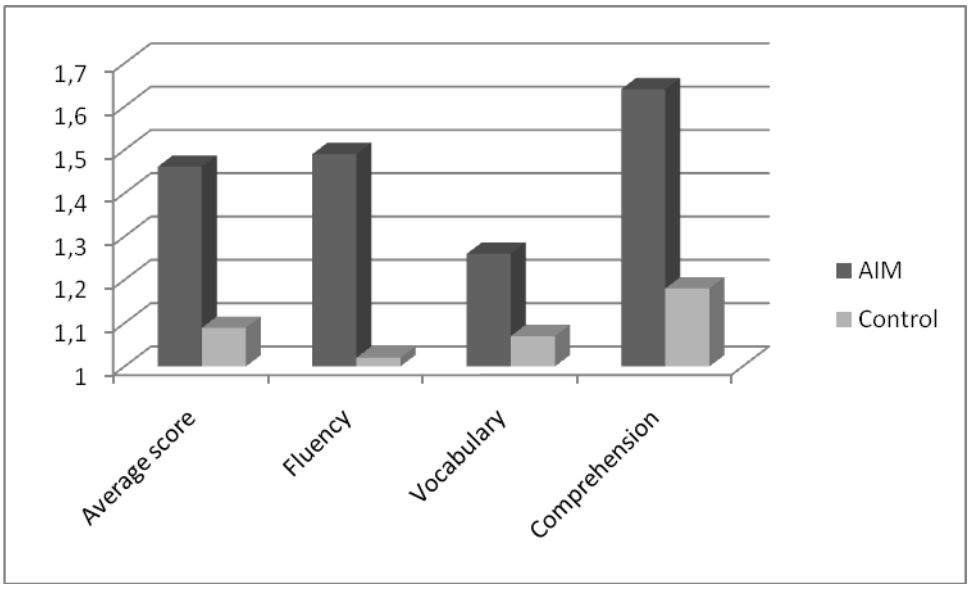

Figure 6. Histogram of the difference in mean score of the two groups on all competences

The impact of intelligence

In previous parts of this paper, we have seen that the teaching methods involved in AIM differ considerably from those of Carte Orange. The more communicative based instruction on which AIM relies might suggest that the methods impacts differently on students.

In second language learning, as in all other sorts of learning, individual differences exist between learners. The one learner will progress more quickly than the other, which can be due to a number of factors. One very important factor determining the success of the learning process is intelligence, or, as it's more frequently called in linguistics, aptitude. Though hard to define and/or measure, the presence of an intra-personal factors that determine the ability to learn is undoubted and supported by a great number of studies (Robinson, 2005; Skehan, 2002).

The design of AIM focuses mainly on communicative activities, encouraging students to learn French in a natural way rather than in the traditional classroom setting where books are followed and teaching concentrates mostly on written language production and perception. Different students use different learning strategies, and some strategies will suit one student very well, whilst another does not seem to take any profit from that same way of learning.

Some of the teachers we have spoken indicated they expected AIM would achieve better results on the students with lower intelligence. It is sometimes suggested in non-academical settings, that students with high intelligence prefer a theoretical approach to learning whilst students with lower intelligence would benefit more from practical approaches. Although this statement is rude by nature, and students should not be classified, generalized or judged in this way, it is true that different types of students respond differently to different teaching methods (Robinson, 2005). Then again, AIM certainly has a specifically practical approach to language learning and as a result, some people may expect some students to benefit more from this method, whilst others - who prefer theoretical learning ways - would benefit more from traditional methods.

Investigating if the impact of AIM on language learning depends on the type of student is very interesting, but difficult to do at the same time. This is due to the fact that no 'type of student' exists or can be defined. Therefore, one can only look at one characteristic at once.

At the start of the AIM project, one of the doubts on the schools' side was whether AIM could be implemented in their heterogeneous classes (classes of which the students have not been selected by Cito-scores, or other intelligence measurements), because it would be unfair to choose a method of learning French that would only suit particular students, ignoring the rest of the class. Hence the idea of researching the impact of intelligence on the success of different methods of teaching French.

Intelligence or aptitude are hard to measure, but for this research we were provided with the cito-scores of all students, as already explained in the methods section. This is regarded as a reliable measurement of intelligence, and as such, it is valuable data to this study.

In this study, we have already brought forward a difference between students of AIM and student of the control group. In order to investigate if intelligence impacts on the success of AIM the gathered the following data, separating the initial groups into two groups: One group containing half of the students with the lowest cito score, and one group with the students with the highest cito score.

	Cito	Mean	SD	N
AIM	low	1,47	0,6	25

	high	1,45	0,38	22
	total	1,46	0,5	47
Control	low	1,07	0,17	20
	high	1,11	0,25	25
	total	1,09	0,22	45
Total	low	1,29	0,5	45
	high	1,27	0,36	47
	total	1,28	0,43	92

Figure 7. Descriptives of cito versus interview ratings

Because of the three variables involved and our interest in the interactionjection of the type of instruction and the cito score (low or high) we used a univariate analysis of variance to control for a statistically significant difference. This, we could not establish, and therefore we did not find that the effect of the type of instruction was influenced by the cito score. In other words, the students that had received AIM instruction with a high cito score did not benefit more from this method than the ones with a low cito score.

This analysis has not only been performed on the mean scores. For each of the different skills (Fluency, Vocabulary an Comprehension) no significant difference was found.

In conclusion, our research leads to the belief that the success of learning French through AIM in comparison to Carte Orange was not affected by the intelligence or the aptitude of the student.

Teacher's impact

As was made clear in previous parts of this thesis, AIM instruction is very different from the more conventional teaching methods you are likely to find at Dutch High Schools. AIM does not follow the strict guidelines provided by methods such as Carte Orange that provide handbooks and require the teacher to follow (or at least to some extent) the structure provided by the method.

AIM demands a lot more from the teacher, which is something both AIM teachers involved in his research indicated from the start. If teaching AIM really is that different from other teaching methods, this raises the question whether we can assume a teacher to be able to adapt to AIM and obtain good results with that method. AIM's particular nature, involving a lot of gesturing, expressing and other types of non-verbal communication, seems to fit better a teacher with an extravert personality than introvert.

In addition, the effort it takes to learn how to instruct AIM and the preparation for the classes should not be underestimated. It takes days of additional courses or workshops and lots of hours to get an idea of how to make AIM instruction a success, and to master the gestures and other teaching methods involved in AIM. Plus, as repeatedly mentioned by the AIM teachers included in this research, preparing the AIM classes takes more time and more effort, even though this remark was often followed by a convincing "but it's all worth it".

All of these indications justify investigating the influence of the teacher on the results obtained by AIM.

In this research, 4 classes and 2 teachers were involved. Each teacher instructed one class with Carte Orange, and one class with AIM, making this setting ideal for a comparative study examining the influence of the teacher on the teaching method.

The relation between the variables teacher and instruction resulted in the following outcomes:

	Teacher	Mean	SD	N
AIM	A	1,33	0,52	24
	B	1,59	0,46	23
	total	1,46	0,5	47
Control	A	1,17	0,29	22
	B	1,01	0,07	23
	total	1,09	0,22	45
Total	A	1,25	0,43	46
	B	1,3	0,44	46
	total	1,28	0,43	92

Figure 8. Descriptives of teacher vs. instruction

All of the data in the figure above is very interesting, but to our research the mean scores of teachers A and B for both instruction types are most eminent. These data resulted in the figure below, depicting the performance of the AIM and Control groups per teacher, and the relationship between those two variables.

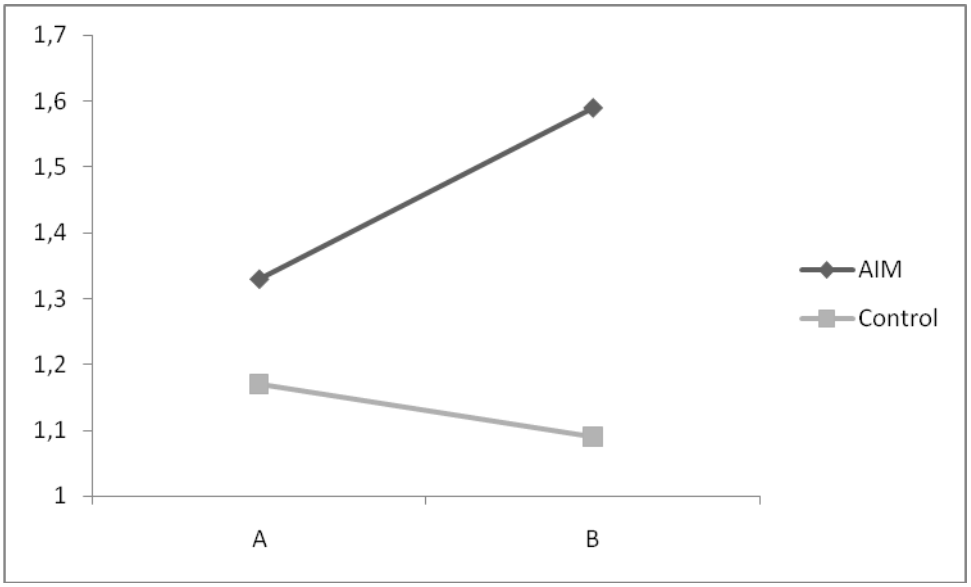

Figure 9. Mean scores of AIM and Control group for teacher A and teacher B.

Both of the figures above suggest the teacher impacts on the effectiveness of the teaching method. Although scores of the AIM students were higher than those of the Control groups for both of the teachers, the difference between the groups is much higher for teacher B (0,5) than for teacher A (0,16). The mean score of the AIM group taught by teacher B obtained, by far, the highest scores. So, AIM instruction was most effective when taught by teacher B. The AIM class of teacher A also outperformed their colleagues of the Control group, but the difference was much lower.

Of course, the differences found were tested for significance:

A Univariate Analysis of Variance (ANOVA) of Between-Subject Effects pointed out that there was an effect in the interception of the variables Instruction (AIM and Control) and Teacher (A and B)on the mean score. This effect was significant: $F(3, 88)= 6{,}805$, $p < 0.05$.

At the same time, the Control group of teacher A performed better than the Control group of teacher B. This means two things.

Firstly, this excludes the conclusion that the high results of the AIM group of teacher B are due to the simple fact that teacher B is a better overall teacher than teacher A. Secondly, and consequently, the data above implies that teacher A is a better Control group teacher, whilst teacher B is the better AIM teacher.

Qualitative results

Initially, the objective of the interviews was to obtain reliable ratings on a set scale for the three skills of Oral Fluency, Oral Vocabulary and Listening Comprehension. During the interviews however, we quickly became aware that there was much more to retrieve from the interviews than just these quantitative assessments. Numerous differences between the AIM and the Control group students could be identified during these interviews, which proved to be a very rich source of information about the effects that two different teaching methods could have on the Oral Production skills of a set of students.

Unfortunately, the rating sheets that we used for the interviews did not provide the means to perform a valid research on detailed aspects of Oral Language Production such as, by example, phonology, and as said, this was not part of the initial research goals either. At the same time, from a more realistic perspective, there is only so much you can do during an interview, and no matter your effectiveness and productivity, there will probably always be something more you would have wanted to do in hindsight.

However, the remarks and conclusions that we have drawn from the interviews regarding the difference between the AIM students and the Control group students are valuable and not mentioning them would be a waste. They assist further interpreting of the results and could inspire future research. Thus, they are relevant to this study, which is why we will summarize these findings in this paragraph.

We realize that the lack of statistical evidence somewhat downgrades these findings, but they are based on the individual opinions and remarks of three different researchers who have unanimously agreed upon these findings. In addition, the notes on the rating sheets taken during the interviews were used in order to formulate these findings, and the video-records of the interviews served as a final verification. Where possible we will give examples from the interviews that support or demonstrate our findings. It is hard to make generalizations based on individual examples, but the examples have been chosen for their stereotypical character of the particular finding we wanted to emphasize.

Probably the most salient different between the AIM students and the Control group students was the nature of the overall conversation. In the case of the Control group students, it was hard to even speak about a conversation, because they were, apart from some exceptions, very one-sided. The interviewer asked questions and gave orders, whilst the students answered or executed the order. That is, given that they understood what they were asked, which often was not the case. The interviewers often needed rephrasing, repetitions and clarifications in order to make themselves comprehensible towards the Control group students. Of course, the AIM students did not understand everything either, but there was

a clear difference between the groups (as demonstrated in the first part of chapter) which allowed more natural like conversations with less interruptions. Most of the AIM students were capable of correctly comprehending and answering simple informal questions and showed willingness to be part of the conversation. Compared to the Control group, the Aim students had a very active role in the conversation.

Also contributing to the unnatural communication with the Control group students was the hesitation most of these students showed towards speaking French. Obviously, this was something they were not used to, which clearly impacted on the conversation. We realize that this is a factor that is highly related to individual personality, and it is true that this hesitation did not present itself in all cases. It is impossible to indicate exactly which emotions played a role in this hesitation, but if it was embarrassment or anxiety, the difference with the AIM students was obvious. A good example is the first thing we started with: 'Ça va?' (How are you?). This very simple question resulted in laughs and looks of embarrassment - once, even the question if they really had to speak French - amongst the Control group students. Whereas the Aim students seemed to like having to talk in French and were comforted by this first easy question to which they knew how to respond. Instead of being startled by having to speak French, the AIM students generally showed a positive, or even enthusiastic, attitude to speaking the target language.

Another interesting aspect of the interviews with the Control group students was the way they responded when they did not understand a question, or when they faced new unknown words or constructions. When this occurred, the Control groups students often responded in Dutch - sometimes even in English - that they had never learned this word. Other students used gestures and mimics to demonstrate their not understanding. The AIM students on the contrary were capable of demanding clarification or expressing their incomprehension in French, because they knew (and used) phrases like 'Je ne comprends pas' (I don't understand), 'Pardon' (pardon me) and 'Qu'est-ce que c'est?' (What's that?).

Besides demonstrating a lack of ability to cope with unknown situations, the phenomena above also indicates that the Control group students had more difficulty separating the languages than the AIM. The AIM students only rarely switched back to Dutch, and never appeared to deal with other interfering languages. The Control group on the other hand consciously and subconsciously mixed languages. These experiences included Dutch, English and even German utterances. Especially when confronted with situations in which they felt unable to express themselves in French, the switch to another language was surprisingly frequent. In addition, the Control group students often miscomprehended words because these words resembled words from another language. The word 'chat' (cat) was taken for the Dutch 'schaap' (sheep) on multiple occasions because of some phonological resemblance and because it fitted the context. Another similar example is the response one of the students gave to the order 'lève-toi' meaning 'stand up': he pointed to his left foot. Apparently, AIM students were capable of speaking French and shutting down other languages, whilst the Control group students still relied on other languages.

Examining the actual utterances of the two groups we also found major differences. Although big differences were present within the groups as well, a clear difference between the groups existed as well. The Control groups students utterances rarely surpassed the level of single isolated words and memorized structures. Questions were almost always reacted to with 'oui' or 'non' and when challenged to build phrases and be creative with their knowledge of the language, they failed to do so. The AIM students performed surprisingly good at this aspect, which may be seen as one of the greatest achievements of AIM. Not only did the AIM students dispose of the means to construct their own phrases, they also tried to do so, sometimes more successfully than at other times. They knew the

different personal pronouns, how to use them, and sometimes even how to conjugate the corresponding verb. It was impressive to watch how the students first chose the pronouns, than the verbs, followed by a noun. Sometimes you could even see them thinking about the sex of the noun and the corresponding article. This strategy allowed them to create their own sentences (though still at a very basic level) and to be creative and express themselves, which is, in the end, what the French educational system strives for.

One last observation we made was that the pronunciation of the AIM students was a lot better than that of the Control group. Of course, this was to be expected. The AIM students first heard the words, before learning to write them, and for the Control group this learning process developed vice versa. Often, the Control group students pronounced words and sounds as if they were Dutch, whilst the Aim students attained a more authentic French phonology.

5. DISCUSSION

The results section of this thesis provided us with a clear overview of the outcomes and the findings of the interviews. It compared the results of the AIM students with those of the Carte Orange students of a set of skills of oral production. Furthermore, the results were compared to different factors, thus investigating the influence on AIM of intelligence and the teacher. The chapter ended with a summary of qualitative findings and impressions resulting from the interviews.

In this chapter, we will be discussing those results, focusing on finding the right interpretations and explanations for the outcomes as demonstrated in the previous chapter. Not only will we look at the results section, we will also refer to the theoretical background and the methods section in order to justify our interpretations.

Interpreting the quantitative results

The interviews brought forward a significant difference in the mean scores on the oral production in the interviews of the two groups. This was to be expected for a number of reasons, of which the most important is undoubtedly the design of AIM instruction versus the design of Carte Orange.

Starting the very first lesson, AIM instruction focuses mainly, sometimes even solely, on developing oral production and oral reception skills. They acquire the linguistics structures by hearing, pronouncing and memorizing gestures. This way of learning is very different from the one forwarded by the Carte Orange method, which emphasizes written production and reception. Even though it aims at improving all skills involved in language development, including oral production and reception, they manifest an evident emphasis on skills needed for written production and comprehension. More on the difference between the methods can be found in the introductory and theoretical background sections. Briefly stated, AIM focuses on oral production whilst Carte Orange tries to develop all skills simultaneously, but with a clear focus on written skills.

Consequently, there is a big difference in the amount of time spent on developing the oral skills between the two methods of instruction. As with every learning process, holds that the more time invested in learning, the better the results. The students of AIM being much more experienced speakers of French, it was to be expected that they would perform better in the interviews than their colleagues who had been taught with Carte Orange.

Another factor that could play a role in these results, is the motivation of the students. As commonly accepted in the field of applied linguistics, motivation and attitude is one of the biggest (if not, the biggest) predictors of success in language development. The main reason for designing AIM was to provide a method that appealed to students, hoping to make the students enthusiastic and motivated towards learning French, which traditional methods had failed to achieve for so many years. Although we did not measure this, we did get the impressions that the groups had a different attitude towards the language, and this could very well have impacted on the results. As said, this has not been measured, but seems to be a very interesting topic for future research.

Of course, a number of others factors can be identified that could explain the results we obtained. The above ones are the most general and important ones. In the paragraphs to come we will present other explanations that are related to the more detailed findings we summarized in the previous chapter.

When we look at the scores for the skills of oral production we distinguished - fluency, vocabulary and comprehension - there are multiple observations we can make.

For both groups holds that the highest scores were obtained for the listening comprehension skill. According to us this is due to our assessment rather than it being the result of a particular aspect of the type of instruction. This trend has been found for both methods and our interpretation of this finding is that the rating scale we used simply provided higher rates for listening comprehension than for the other skills.

This thesis' main objective is to compare two methods of instruction and therefore any deviations are much more interesting to us. As such, the lowest scores of the AIM group were obtained, by far, for oral vocabulary, whilst the Carte Orange students performed least good on Oral Fluency. Again these results can be explained by looking at the design of the methods.

AIM instruction, as already stated, present a very practical approach to language learning focusing on the students´ ability to express themselves in the target language. Out of the three skills, oral fluency is the one that relates most to this focus, thus explaining why they performed relatively good on this skill. We consider oral vocabulary to be the skill that relates least good to the instruction forwarded by AIM, because it does not attribute as much time and energy to explicitly acquire words in the target language as other traditional methods such as Carte Orange. One of the theories inserted in the design of AIM is to treat less words, but to repeat them more often. The end goal is to make sure more words are actually acquired, rather than learned once and then forgotten. This method proved effective, because the AIM students were rated statistically significantly higher on vocabulary than the Carte Orange students. Still, the goal of learning less words, but learning them better, explains why the AIM students performed less good on vocabulary than on fluency and comprehension.

As opposed to the AIM students, the Carte Orange students performed relatively well on vocabulary, and a lot less well on fluency. This is due to the fact that they were not very experienced speakers of French and not very able to converse in French. We will come back to this aspect later. However, the Carte Orange students did sometimes surprise us with their knowledge of vocabulary. Out of the blue, they would present words we did not expect them to know because it did not suit the level of French they had attained before. In addition, the Carte Orange were better counters in French than the AIM students. We interpret these aspects as being the effect of explicit learning of lists of words, an aspect that AOM does not know, whilst very frequent in Carte Orange. The effect: The students had acquired some low frequency words, whilst not having mastered some of the most simple and frequent words and structures.

The results did not suggest in any way that the success of AIM depends on the intelligence of the students, which is a very strong point of AIM. Apparently, the method is capable of achieving good results, no matter the intelligence of the students.

However, this finding can be considered surprising at the same time. AIM was designed for a number of reasons, one of them being to provide a method that was less theoretical and more suitable for all types of students. It is often argued that some students require a different kind of instruction with other learning strategies than other students. Some students learn a word most quickly by hearing its pronunciation, others will prefer to see it written on a paper in order to take it in. Many researchers have even argued about the existence of multiple kinds of intelligence (Gardner, 1985) that all have a different impact on learning processes. AIM was designed in such a way that it tries to stimulate as much as possible all of these different intelligences. For example, the gestures stimulate the, as Gardner calls it, *bodily-kinaesthetic intelligence*. Students that score high on this type of intelligence are likely to benefit

more from AIM than from methods that do not include these gestures. As opposed to Carte Orange, AIM strongly stimulates other intelligences as the *interpersonal intelligence* (through peer communication) and the *musical intelligence* (think of the high amount of music and theatre included in the method) as defined by Gardner. These intelligences are highly related to communicative and practical activities. Methods as Carte Orange aim more on stimulating the *linguistic*- and *logical-mathematical intelligence* which are more theoretical by nature.

Without starting a discussion about the definition of intelligence, it is important to make this distinction. Roughly looking at it, the cito score we used in this study cannot be considered a measure of, for example, *musical intelligence*. From Gardners perspective, it is mostly a measure of *linguistic*-, *logical-mathematical*-, and *spatial intelligence*. Other intelligences, are only slightly included in the test. Since these intelligences suit better the design of Carte Orange, you would expect the cito test to be a better predictor of learning success for the control group students, than for the AIM students. However, on average, for both groups holds that a higher cito score resulted in higher ratings on the interview. Apparently, the cito score was a good predictor of learning success and hence a good indicator of the ability to learn a language.

It is often argued that students that have a high overall intelligence benefit more from theoretical approaches to language learning than students with lower overall intelligence, who, on the contrary, prefer a more practical communicative approach. From that perspective, looking at the designs of our methods, one would expect to find that AIM impacts differently on the various students. The results show that no matter the intelligence, all students benefit from AIM. However, the above suggest that the students with high intelligence could be expected to benefit less from AIM than those with lower intelligence scores (and the other way around), because they prefer more theoretical and less practical and communicative learning strategies. This effect was not found.

A definite explanation for not finding this effect we cannot provide. Apparently, the success of AIM is not affected by the intelligence of the students, which is a very strong point of the method. On average, better results were obtained no matter the cito score of the students, where the design of the method aims at appealing and stimulating the students with less theoretical learning strategies in particular. These results imply that AIM's effectiveness does not depend on intelligence, which renders it applicable in the different layers of education.

We already quoted one of the teachers of Aim at the Werkman College once before in this thesis, stating that AIM instruction demands a lot from a teacher, but it's worth it. At that time, we wanted to underline the worth of AIM for teachers, but now we insist on the first part of that quote, being that AIM instruction is a lot of work.

This statement that AIM takes more effort from a teacher than previous methods is something that has been confirmed by all teachers involved in this research. In order to learn to instruct French with AIM teachers follow courses and need self-study to learn the gestures, objectives and important theory behind the method. The preparation before each lesson is more demanding as well. AIM does not provide a structured chronology of lessons and exercises as do other methods as Carte Orange. In AIM instruction, teachers have more freedom to construct their planning and their lessons as they think is best. Although this may seem to be a very positive aspect of AIM, and we think it is, but it does require the teacher to think about and prepare each and every lesson. Therefore teachers have to be able and ready to invest both time and effort into their lessons.

Not only the preparation of the instruction takes effort, the lessons themselves have proven to be demanding as well. As the teachers indicate, AIM lessons are more exigent than other lessons. AIM

requires the teacher to constantly be highly active (mentally rather than physically) throughout the lessons. The teacher needs to speak relatively much and accompany that speech with the gestures. In addition it is very important for the teacher to have a positive attitude, in order to keep the students motivated and activated. For example, when you play and sing a song together with the class, the teacher will need to encourage the students to sing along, which is very hard if he/she does not like to sing the song him-/herself. If the teacher is not positive about the song, how can he/she expect the students to be. Same story for the theater plays and the storytelling that are heavily involved in AIM. This also explains why both the AIM instructors and the teachers at the Werkman point out that teachers should not be too introverted in order to be able to use AIM, because this will result in a negative attitude of both teacher and student. As one of the teachers stated, teaching with AIM can be very fulfilling and give you a positive vibe, but if you are having a bad day, teaching AIM can be difficult. She stated it was hard to keep smiling and keep your enthusiasm towards the lessons and the students.

In addition to AIM being highly time and effort consuming, it seems to demand some personal characteristics as well. These are two factors that impact on the results of AIM, and hence, may imply that not every teacher may be equally suitable as a teacher of AIM, or will obtain the same results following AIM. The comparative results section treating the influence of the teacher on the effect of AIM confirms this implication. The teacher factor is one that is decisive for the effectiveness of the lesson of AIM. Both teachers however obtained better results with their AIM group than with the Control group. We can only guess about the reasons of the difference between the teachers, but we believe the personality of the teachers is of significant importance when teaching with AIM. This however has not been tested, and therefore opens a door to future researchers interested in this topic.

Interpreting the qualitative results

In the previous chapter, we mentioned that the character of the conversation probably was the most salient difference between the interviews with the AIM students and the control group students. Where we could hardly speak of an actual conversation with the control group students, the AIM students were more active and more capable of participating in ongoing informal conversations. We consider two factors to be the major underlying causes that explain this finding.

Firstly, the AIM students were far more experienced speakers of French. Oral production is a large component of AIM instruction, while Carte Orange focuses mainly on written production and comprehension. Each AIM lesson the teachers starts by asking how they are doing, what the weather is like, etc. As a consequent, the AIM students are very familiar with the informal questions and answers, and the fact that they know better how to handle themselves in an informal conversation than their colleagues from the control group was to be expected.

Secondly, in general the AIM students used a different strategy for expressing themselves. As mentioned in the previous chapter, the control group relied mostly on isolated words, as opposed to the AIM group, that tried to construct phrases with the structures they had acquired. Often they followed the (correct) SVO structure, starting with a pronoun, followed by a conjugated verb and ending the phrase with a noun. This strategy allowed them to be creative with the language, and thus express themselves. Though sometimes lacking the vocabulary to do so correctly, they surpassed the level of repeating structures they had learned by heart. The ability to express themselves and to be creative with the language contributed strongly to the quality of the conversation.

But where does this ability to construct phrases of the AIM students come from? We believe that, again, the answer to this question can be found in the design of AIM instruction. One of the constructs imbedded in the AIM is to begin by learning only the words and structures that are most frequently used

and provide the means to communicate. It does not instruct as many words as other methods (as for example Carte Orange) but the words that are highly frequent and as such very valuable building stones for producing French. The control group students may have surprised us at times by knowing a relatively difficult word, they have surprised us even more by not knowing the simplest of words and constructions, disabling them to communicate. Because the AIM focuses on learning these basic, but very important, words and structures, the AIM students had learned these highly frequent words better and were able to use them in ongoing speech.

Another difference between the groups we mentioned previously is the high level of activation of other languages during the interview we found in the control group, whereas the AIM group managed a lot better to keep speaking the target language. We know that the level of activation is correlated to the L2 proficiency. Dijkstra, Timmermans and Schriefers (2000) have shown us that the effect of the L2 on the L1 is weaker when the proficiency level in the L2 is low, and vice versa. In other words, the lower the L2 proficiency, the higher the effect of the L1 on that L2. The quantitative results have given us reason to believe that the AIM students indeed have a higher proficiency of French than the Carte Orange students, thus explaining why the L1 (and eventual L2 or even L3) show higher levels of activation in the control group.

We also found that the AIM students showed a better French pronunciation. The AIM students learned words by hearing them before learning to write them, whereas the Carte Orange students learned words from another approach. They had to be capable of recognizing and writing the word, rather than knowing how to pronounce it. Consequently, the control group students often came up with typically Dutch pronunciations of French words. They lacked awareness of the French phonology, because they did not need a good pronunciation. The AIM student on the contrary were forced to speak French during the lessons, so in order to make themselves comprehensible they needed at least some phonological knowledge of French.

Interpreting human emotions is hard to do in real-life, and even harder if they need to be academically acceptable. Therefore it is hard to explain the anxiety or embarrassment towards speaking French we experienced in so many students in the control group, and of course we cannot assume all students to have felt the same for the same underlying reasons. Still, it was obvious that the control group students were less at ease speaking French than the AIM student, and all of the factors named above probably played a role in this given. Summarizing this paragraph in one single sentence, the control group students were less experienced and less capable speakers of French, of which they were probably aware. So, not only were they asked to do something they were not very experienced in, they also realized they were not very good at it. These are two factors that have probably made the control group students less confident than the AIM students, who were more experienced and capable.

In the next chapter we will briefly go over all of the previous chapters and emphasize the most important results, findings and interpretations. In addition, we will also express our feelings about AIM and its future. At the same time, we will indicate where future research is necessary or could be of value.

6. CONCLUSION

This last chapter will provide an overview of the previous chapter, whilst it also contains our answers to the research questions we formulated in the theoretical background section. Moreover, it is the place where we express our feelings and impressions about AIM and bring up some ideas, which could inspire other researchers interested in this topic.

The first and main research question was about whether or not the students who have been taught by AIM have better oral language production skills than those who have been taught by Carte Orange. The quickest and simplest of answers our study could give is: Yes.

The interviews, in which we assessed the students' oral proficiency provided a significant difference between the superior AIM students (mean= 1,461) over the students of Carte Orange (mean= 1,089). This averaged score included the three ratings the students had been awarded on the three oral communication skills of oral fluency, oral vocabulary and listening comprehension. On each of these skills (as demonstrates figure 10) the AIM students outperformed the Carte Orange students. These differences were significant.

	Fluency	Vocabulary	Comprehension
AIM	1,49	1,26	1,64
Carte Orange	1,02	1,07	1,18

Figure 10. Mean scores per group per oral communication skill

The raw numbers obviously speak in favor of AIM over Carte Orange when it comes to oral language production, but do the qualitative results from the interviews.

From our observations we could deduce that the AIM students were more capable of having an informal conversation in French, than their colleagues who had received Carte Orange instruction. They were less hesitant or embarrassed to speak, and proved themselves more experienced speakers. In addition they showed the willingness to construct phrases by combining structures they had acquired, which was very impressive compared to the (mostly) isolated words and expressions the Carte Orange students used. Moreover, the AIM students were better capable of separating languages and staying in the target language, and appeared to have a slightly better pronunciation.

We believe that most of these differences are due to the design of AIM, that includes aspects from which the students appear to benefit and attributes a lot more attention to developing oral production skills than Carte Orange. Even though we did expect to find a difference, we did not expect it to be this convincing. We sincerely hope that the progress the AIM students have made during their first year will continue. Due to the design of AIM, including the music, stories and drama, it seems to be very suitable to students of younger ages. It is very interesting to see how AIM as a method, and how the students will develop in the coming years. We know other projects have already been started, so more on the endurance of the positive AIM effects will hopefully come soon.

The answer to our second research question, about the influence of scholastic aptitude on the effectiveness of AIM, is less conclusive. We did find that scholastic aptitude impacted on the results, but we did not find that the effect of AIM was determined by scholastic aptitude.

This somewhat surprised us, because AIM's musical and theatrical aspects are so dominant and practical that the method would seem to appeal more to the extravert and communicative students. The students with a high score on scholastic aptitude would maybe prefer a more theoretical approach. However, we did not find this effect in our study.

An effect was found whilst investigating the teacher variable. We found a strong significant difference between the results of the two teachers that were involved in this research. One of the teachers performed a lot better with her AIM class than the other, who performed considerably better with her Carte Orange class. This difference leads to believe that the effectiveness of AIM depends on the teacher, because AIM apparently demands other capacities or characteristics of a teacher than Carte Orange. Even if we may have measured a significant different, it has to be noted that this study only involved two teachers, and no further research has been executed on the kind of characteristics that could explain these differences. It is however important to be aware of the possibility does not suit evenly well to all teachers, which makes the method harder to implement universally.

In addition to answering the research questions, we would also like to use this chapter to express some of our ideas and impressions about the method. These are based on the experiences we have had during this research. As very often, each new insight raises new questions. These ideas can also be considered as the questions about AIM that have arisen in our minds throughout this study. AIM has shown itself to be a very promising method, but we do think there is a lot more to explore about this method. Thus, we hope these ideas may inspire future research.

AIM as a method of teaching French is recent and still in development. The very playful design appears effective during the first years, but as the students age will have to be adjusted. The big question is if AIM can continue the good results it obtains in the first year, and if not, how can we ensure that it does. Moreover, implementing AIM poses practical issues, because it simply demands more from both school and teachers. Plus, one of the problems the H.N. Werkman will face in a few years is what will happen when the classes start to mix up. Is it possible to have former Carte Orange students in the same class as former AIM students, and how will they develop in such a case? AIM's popularity continues to build, but there exists insecurity about how the method will affect students on the long term.

AIM's popularity increases explosively and it is not surprising that the method has drawn attention of the other languages that are taught in High School as well. It is very plausible that people are already working on similar methods for German, English and/or Spanish. We have thought about this and question whether this would be a positive development. We think that one of the reasons why AIM works is because it is different from other methods. It is salient and appeals to the students, who form a connection between AIM as a method and French as a language. We fear that if students are taught different languages following this same method, this will decrease the positive effects of AIM. Obviously, further research on this issue is necessary and valuable.

A last impression we would like to express is that the process of SLA that AIM follows resembles more closely than Carte Orange (and other similar methods) the L1 way of learning. Learning the language is done by hearing and speaking first, ignoring written language at first and without any explicit rules of

grammar or vocabulary. This observation may seem rudimentary, but if this is indeed the case, it may provide us better insights in the processes in which AIM is involved and could help us predict its future results and improve its design.

In conclusion, this thesis provides a conclusive answer to the question if AIM has a positive effect on oral language production for our sample groups. The answers to the other research questions are less definitive, but open ways to future research. AIM is a very recent development within the educational domain, and due to its popularity and the explosive increase of schools that use AIM it is important that more academic data is gathered on the effects of AIM. How AIM will develop in the coming years is to be seen, but the first signs are very promising. Hopefully, this research constitutes a stimulation for studies on AIM and can hopefully inspire other researchers.

Daan Jans

Wordcount:18.292

BIBLIOGRAPHY

Allen, J. (1989). Reading and writing development. In J. Mason (Ed.), *Reading and writing connections*, Needham Heights, MA: Allyn & Bacon.

Arnott, S. (2005). The accelerative integrated study: A descriptive case study. Unpublished qualifying research paper, Ontario Institute for Studies in Education, University of Toronto.

Barry, C., & Seymour, P. H. K. (1988). Lexical priming and sound-to-spelling contingency effects in nonword spelling. *Quarterly Journal of Experimental Psychology*, *40A*, 5–40.

Boyson, B.A., Rhodes, N.C., & Thompson, L.E. (2009). Administrator's manual for CAL Foreign Language Assessments, Grades K-8. Center for Applied Linguistics, Washington, DC.

Bourdages, J., & Vignola, M.J. (2009). AIM : La communication orale chez les élèves de l'élémentaire en français de base. Canadian Modern Language Review, 65(5), 731–756.

Burstall, C. (1968). French from eight: A national experiment. Windsor, UK: National Foundation for Educational Research.

Burstall, C. (1970). French in the primary school: Attitudes and achievement. Windsor, UK: National Foundation for Educational Research.

Chomsky, N. (1966). Research on language learning and linguistics. In Report of the Northeastern Conference, 1966.

Cummins, J. (1996). Negotiating Identities. Toronto: California Association for Bilingual Education.

Dijkstra, A.F.J., Timmermans, M.P.H., & Schriefers, H.J. (2000). On being blinded by your other language: Effects of task demands on interlingual homograph recognition. *Journal of Memory and Language, 42* (4), 445-464.

Gardner, H. (1985). Frames of Mind: The Theory of Multiple Intelligences. New York: Basic Books.

Harley, B., Swain, M. (1984). The interlanguage of immersion students and its implication for second language teaching. In Davies, A., Criper, C., & Howatt, A.P.R. (Eds.), *Interlanguage* (pp. 291-311). Edinburgh, Scotland: Edinburgh University Press.

Harley, B., Lapkin, S., Scane, J., Hart, D., & Trépanier, C. (1988). Testing outcomes in core French: The development of communicative instruments for curriculum evaluation and research. Toronto, ON: Modern Language Centre, Ontario Institute for Studies in Education, University of Toronto.

Krashen, S. (1981). Second Language Acquisition and Second Language Learning. New York: Pergamon Press.

Liskin-Gasparro, J.E. (1983). Teaching and Testing Oral Skills. Presidio of Monterey, California.

Long, M.H. (1991). Focus on form: a design feature in language teaching methodology. In K. De Bot, R. Ginsberg, and C. Kramsch (Eds), *Foreign language research in cross-cultural perspective* (pp. 40-52). Amsterdam: John Benjamins.

Mady, C., Arnott, S., & Lapkin, S. (2007). A comparison of AIM and non-AIM Grade 8 core French in the Bluewater District School Board: Students' French proficiency and teacher and student perceptions. Unpublished report, Modern Language Centre, Ontario Institute for Studies in Education, University of Toronto.

Mady, C. Arnott, S. & Lapkin, S. (2009). Assessing AIM: A study of Grade 8 students in an Ontario school board. *Canadian Modern Language Review, 65* (5), 703-729.

Maxwell, W. (2001). Evaluating the Effectiveness of the Accelerative Integrated Method for Teaching French as a Second Language. Unpublished master's thesis, University of Toronto: Ontario.

Menezes, V. (2009). Second language acquisition: from main theories to complexity (reelaboração do trabalho apresentado no congresso da AILA 2008).

Michels, M. (2008). Innovation in French as a second language teaching at the elementary level:

an exploratory investigation (Unpublished master's thesis). University of London Institute in Paris, France.

Piaget, J. (1956). The origins of intelligence in children. (M. Cook, Trans.). NY: International University Press.

Robinson, P. (2005). Aptitude and Second Language Acquisition. *Annual Review of Applied Linguistics, 25,* pp. 46-73. Cambridge University press.

Schwarz, B.D. (1993). On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition*, *15*, 147-163.

Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition*, 15, 165-179.

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69–93). Amsterdam: Benjamins.

Spada, N., & Frohlich, M. (1995). COLT Communicative Orientation of Language Teaching Observation Scheme Coding Conventions and Applications. Sydney: National Centre for English Language Teaching and Research (NCELTR).

Turnbull, M. (1999). Multidimensional Project-based Second Language teaching: Observations of

Four Grade 9 Core French Teachers. *The Canadian Modern Language Review, 56,* 1,7-29.

Wesche, M. (1994). Input and interaction in second language acquisition. In C. Gallaway & B. Richards (eds) *Input and Interaction in Language Acquisition* (pp. 219-249). Cambridge: Cambridge University Press.

APPENDICES

APPENDIX I RATING SHEET

The SOPA rating sheet as provided by the Center for Applied Linguistics, copied from Boyson, Rhodes, & Thompson, (2009):

CAL Center for Applied Linguistics

Appendix D

Student Oral Proficiency Assessment (SOPA) Rating Sheet

Language of Interview: ☐ Arabic ☐ Chinese ☐ English ☐ French ☐ German ☐ Japanese ☐ Russian ☐ Spanish ☐ Other (check one box)

School _____ Grade _____ City/State_____ Interviewer _____ Rater _____ Date _____

Student Information	Task 1( )	Task 2( )	Task 3( )	Task 4( )	Ratings
Student A Name: Language background:					Oral Fluency: Grammar: *(Speaking)* Vocabulary: *(Speaking)* Listening Comprehension:
Student B Name: Language background:					Oral Fluency: Grammar: *(Speaking)* Vocabulary: *(Speaking)* Listening Comprehension:

Additional Comments:

APPENDIX II RATING FORMAT

The SOPA rating format as provided by the Center for Applied Linguistics, copied from Boyson, Rhodes, & Thompson, (2009):

CAL CENTER FOR APPLIED LINGUISTICS ▪ 4646 40th Street, NW ▪ Washington, DC 20016-1859 ▪ http://www.cal.org

Appendix B

CAL ORAL PROFICIENCY EXAM AND STUDENT ORAL PROFICIENCY ASSESSMENT RATING SCALE (COPE/SOPA-RS)*

English Version © 2009** CAL

JR. NOVICE-LOW	JR. NOVICE-MID	JR. NOVICE-HIGH	JR. INTERMEDIATE-LOW	JR. INTERMEDIATE-MID	JR. INTERMEDIATE-HIGH	Jr. Advanced-Low	JR. ADVANCED-MID	JR. ADVANCED-HIGH
Oral Fluency								
-Produces only isolated words (i.e., single-word responses) and/or greetings and polite expressions such as good morning and thank you.	-In addition to isolated words, uses phrases of two or more words, and/or memorized phrases or sentences (e.g., My name is..., I don't know) in predictable topic areas. -May attempt to create sentences, but is not successful. -Long pauses are common.	-Uses memorized expressions with reasonable ease. -Shows emerging signs of creating with the language to communicate ideas. -Creates some sentences successfully, but cannot sustain sentence-level speech.	-Goes beyond memorized expressions to maintain simple conversations at the sentence level by creating with the language, although in a restrictive and reactive manner. -Handles a limited number of everyday social and academic interactions.	-Maintains simple sentence-level conversations. May initiate talk spontaneously without relying on questions or prompts. -Gives simple descriptions successfully. -May attempt longer, more complex sentences. Few, if any, connectors are used.	-Initiates and sustains conversations by using language creatively. -Shows emerging evidence of paragraph-like speech with some connected sentences (e.g., then, so, that, etc.) in descriptions and simple narratives, but has no actual paragraphs with a main idea, organization, and connection.	-Reports facts easily. Can discuss topics of personal interest and some academic topics at the paragraph level to satisfy school and everyday requirements. -Narrates and describes at the paragraph level also, although haltingly at times. -False starts are common.	-Handles with ease and confidence concrete topics of personal and general interest and a number of academic topics. -Narrates and describes smoothly in paragraphs having a main idea, organization, and a variety of sentence connectors (e.g., first, next, finally; then, when, that, although, but, therefore, so, etc.).	-Handles most social and academic requirements confidently, but may hesitate when responding to complex, formal tasks (Superior level). -Organizes and extends discourse (multiple paragraphs) in an emerging ability to hypothesize on abstract topics (if-then) and support opinions.
Grammar *(Speaking)*								
-May use greetings and polite expressions accurately. -Lacks an awareness of grammar and syntax.	-Memorized expressions with verbs and other short phrases may be accurate, but inaccuracies are common. -Does not successfully create at the sentence level with conjugated verbs.	-Creates some sentences with conjugated verbs, but in other attempts to create sentences, verbs may be lacking or are not conjugated. -Other grammatical inaccuracies are present.	-Uses a variety of common verbs in present tense (conjugations may be inaccurate) in sentences. -Other verb tenses/forms may appear in memorized language. -The listener may be confused by this speech due to the many grammatical inaccuracies.	-Uses an increasing number and variety of verbs. -Verbs are mostly in present tense although awareness of other verb tenses (future/past) and forms may be evident. -Many grammatical inaccuracies may be present.	-Uses a large variety of verbs well in present tense. Uses many verbs in the past tenses but lacks control of past. May use future and other verb forms. -Grammatical inaccuracies may still be present. Awareness of inaccuracies may be evident.	-Uses present, past, and future tenses. -May effectively self-correct when aware of grammatical inaccuracies. -Structures of native language may be evident (e.g., literal translation).	-Has good control of present, past, and future tenses. -Some inaccuracies may remain, but speech is readily understood by native speakers of the language. ***In some cases, may use non-standard varieties of grammar.	-Uses all verb tenses accurately and sometimes uses complex grammatical structures, (e.g., if...occurred, then...might also happen). -Some patterns of error may persist, but they do not interfere with communication.
Vocabulary *(Speaking)*								
-Uses single words in very specific topic areas in predictable contexts. -May use greetings and polite expressions.	- Uses single words, short phrases, greetings, polite expressions, and other memorized expressions on a limited number of topics. -Frequent searches for words are common. May use native language or gestures when attempting to create with language.	-Uses vocabulary centering on basic objects, places, and common kinship terms, adequate for minimally elaborating utterances in predictable topic areas. -Use of native language and gestures is common to expand topics.	-Has basic vocabulary for making statements and asking questions to satisfy basic social and academic needs, but not for explaining or elaborating on them. -Use of some native language is common when vocabulary is lacking.	-Has basic vocabulary, permitting discussions of a personal nature and limited academic topics. Serious gaps exist for discussing topics of general interest. -If precise word is lacking, may use circumlocution ineffectively. May resort to native language.	-Has a broad enough vocabulary for discussing simple social and academic topics in generalities, but lacks detail. -Sometimes achieves successful circumlocution when precise word is lacking. May use native language occasionally.	-Vocabulary is primarily generic but is adequate for discussing concrete or factual topics of a personal nature, topics of general interest, and academic subjects. -May use circumlocution successfully when specific terms are lacking.	-Has adequate vocabulary for including detail when talking about concrete or factual topics of a personal nature, topics of general interest, and academic subjects. -Uses circumlocution effectively. Rarely uses native language.	-Uses precise vocabulary for discussing a wide variety of topics related to everyday social and academic situations. -Lack of vocabulary rarely interrupts the flow of speech.
Listening Comprehension								
-Recognizes single, isolated words, greetings and polite expressions.	-Understands predictable questions, statements, and commands in familiar topic areas (with strong contextual support), though at slower than normal rate of speech and/or with repetitions.	-Understands simple questions, statements, and commands in familiar topic areas, and some new sentences with strong contextual support. -May require repetition, slower speech, or rephrasing.	-Understands familiar and new sentence-level questions and commands in a limited number of content areas with strong contextual support for unfamiliar topics. -Follows conversation at a fairly normal rate.	-Understands sentence-level speech in new contexts at a normal rate of speech although slow-downs may be necessary for unfamiliar topics. -Carries out commands without prompting.	-Understands longer stretches of connected speech on a number of topics at a normal rate of speech. -Seldom has problems comprehending everyday topics. (Can request clarification verbally.)	-Understands main ideas and many details in connected speech on some academic topics and on topics of personal interest.	-Understands main ideas and most details in connected speech on a variety of topics, but may be unable to follow complicated speech. -May have difficulty with highly idiomatic speech.	-Understands complex academic discourse and highly idiomatic speech in conversation. -Confusion may occur due to socio-cultural nuances or unfamiliar topics.

*The *COPE/SOPA Rating Scale* is based on the *ACTFL Speaking and Listening Proficiency Guidelines*, American Council on the Teaching of Foreign Languages (1986, 1999).

**Please note that the 2009 version of this rubric is a revision in terminology only to make the descriptions clearer, and not a revision of the sublevels.

***This feature may not appear, but if present in student speech, is acceptable at the Jr. Advanced-Mid level of proficiency.